A Bayesian-motivated test for linear model in high-dimensional setting

Rui Wang[1] and Xingzhong Xu[*1,2]

[1] School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081,China

[2] Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081,China

Friday 4th January, 2019

Abstract

Using the idea of Bayesian factor, a new test for linear model in high-dimensional setting is proposed.

Our theory is also useful in.

1 Introduction

Consider the high-dimensional linear regression model of the form

$$\mathbf{y} = \mathbf{X}_a\boldsymbol{\beta}_a + \mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^n$ is the response, $\mathbf{X}_a$, $\mathbf{X}_b$ are $n \times q$ and $n \times p$ design matrices, respectively, $\boldsymbol{\beta}_a \in \mathbb{R}^q$, $\boldsymbol{\beta}_b \in \mathbb{R}^p$ are unknown regression coefficients, and $\boldsymbol{\varepsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ are the iid errors with mean 0 and covariance $\sigma^2 = \phi^{-1}$. Here we break the predictors into two parts $\mathbf{X}_a$ and $\mathbf{X}_b$ such that $\mathbf{X}_a$ contains the predictors that are known to have effect on the response, and we would like to know if $\mathbf{X}_b$ contains useful predictors. That is, we are interested in testing the hypotheses

$$\mathcal{H}_0 : \boldsymbol{\beta}_b = 0, \quad \text{v.s.} \quad \mathcal{H}_1 : \boldsymbol{\beta}_b \neq 0. \tag{2}$$

Motivated by many recent applications of high dimensional regression, we consider the situation where $p + q$ is much larger than n. Unless mentioned otherwise, we consider fixed design in this article.

*Corresponding author

Email address: xuxz@bit.edu.cn

The conventional test for hypotheses (2) is the F-test which is also the likelihood ratio test under normality. However, the F-test is not well defined in high dimensional setting. In fact, if $\boldsymbol{\varepsilon}$ is normal distributed and $\text{Rank}[\mathbf{X}_a; \mathbf{X}_b] = n$, then the likelihood is unbounded under the alternative hypothesis. This calls for new test methodologies in high-dimensional setting.

Two different high-dimensional settings have been extensively considered in the literature. One is the small p, large q setting. An important example of this setting is testing individual coefficients of a high-dimensional regression. See Buhlmann (2013), Zhang and Zhang (2014) and Lan et al. (2016) for testing procedures in this setting. In this paper, however, we focus on the other setting, namely the large p, small q setting. In this case, there are just a few covariates, namely $\mathbf{X}_a$, are known to have effect on the response, while there remain a large number of covariates, namely $\mathbf{X}_b$, to be tested. In practice, which covariates belong to the part $\mathbf{X}_a$ is determined apriori. If no prior knowledge is available, $\mathbf{X}_a$ can be $\mathbf{1}_n$.

Many test procedures have been proposed in the large p, small q setting. Based on an empirical Bayes model, Goeman et al. (2006) and Goeman et al. (2011) proposed a score test statistic as well as a method to determine the critical value of their test statistic. Later, Lan et al. (2014) proposed a similar test, but they used asymptotic normality to determine the critical value. Based on U-statistics, Zhong and Chen (2011) proposed a test for the case where $\mathbf{X}_a = \mathbf{1}_n$. Later, a generalization of this test to the general design matrix $\mathbf{X}_a$ is proposed by Wang and Cui (2015). To accomodate outlying observations and heavy-tailed distributions, Feng et al. (2013) proposed a rank-based test for the entire coefficients. Xu (2016) modified Feng et al. (2013)'s test and proposed a scalar invariant rank-based test. Apart from the afore mentioned tests, there is another line of research utilizing desparsified Lasso estimator; see Xianyang Zhang (2017) and the references therein.

We note that the test procedure in Goeman et al. (2006) and Goeman et al. (2011) is motivated by Bayesian methods but is treated as a frequentist significance test. Bayesian-motivated tests are very suitable for high-dimensional testing problems since the Bayes factors corresponding to proper priors are always well defined, even if the likelihood is unbounded. In the Bayesian literature, many Bayesian tests have been proposed for hypothesis (2) in low-dimensional setting; see Javier Girón et al. (2006); Goddard and Johnson (2016); Zhou and Guan (2018) and the references therein. However, most existing Bayesian tests are designed for small p, small q setting and are not well defined in large p, small q setting. In this paper, we propose a new test statistic in large p, small q setting which is the limit of Bayes factors under normal linear model. We prove that, under mild conditions, the distribution of the proposed test statistic can be accurately approximated using Lindeberg's replacement trick. And the critical value is determined by this approximated distribution. Under stronger assumptions, we also derive the asymptotic power funcion. A simulation is conducted to examine the performance of the proposed test.

The rest of the paper is organized as follows.

2 Methodology

Testing hypotheses (2) in large p, small q setting is a challenging problem. As Goeman et al. (2006) noticed, if $\boldsymbol{\beta}_b \neq 0$ but $\mathbf{X}_b\boldsymbol{\beta}_b = 0$, no test has any power. They also pointed out their test has negligible power for many alternatives and consequently is not unbiased. For low-dimensional testing problems, a biased test is often regarded as problematic. However, the following proposition shows that under normal assumption, there is no nontrivial unbiased test in large p, small q setting.

Proposition 1. *Suppose* (1) *holds with* $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n)$*. Suppose* $\mathbf{X}_a$ *is an* $n \times q$ *matrix with full column rank,* $q < n$ *and* $\mathrm{Rank}([\mathbf{X}_a; \mathbf{X}_b]) = n$*. Let* $\varphi(\mathbf{y})$ *be a test function of level* α*, that is, a Borel measurable function satisfying* $0 \leq \phi(\mathbf{y}) \leq 1$ *and* $\mathrm{E}[\phi(\mathbf{y})] \leq \alpha$ *under the null hypothesis. If* $\phi(\mathbf{y})$ *is unbiased, that is* $\mathrm{E}[\phi(\mathbf{y})] \geq \alpha$ *under the alternative hypothesis, then* $\varphi(\mathbf{y}) \equiv \alpha$*, a.s.* λ*, where* $\lambda(\cdot)$ *is the Lebesgue measure on* $\mathbb{R}^n$*.*

The above proposition implies that it is impossible to find a test with reasonable power for all alternatives. This motivates us to adopt Bayesian methods to find a test with good average power behavior. Within the Bayesian framework, Bayes factor is commonly used for comparing two models. In our problem, suppose $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n)$, the Bayes factor for hypotheses (2) is

$$B_{10} = \frac{\int d\mathcal{N}_n(\mathbf{X}_a\boldsymbol{\beta}_a + \mathbf{X}_b\boldsymbol{\beta}_b, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\pi_1(\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi)\, \mathrm{d}\boldsymbol{\beta}_b \mathrm{d}\boldsymbol{\beta}_a \mathrm{d}\phi}{\int d\mathcal{N}_n(\mathbf{X}_a\boldsymbol{\beta}_a, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\pi_0(\boldsymbol{\beta}_a, \phi)\, \mathrm{d}\boldsymbol{\beta}_a \mathrm{d}\phi},$$

where $d\mathcal{N}_n(\mu, \boldsymbol{\Sigma})(\mathbf{y})$ is the density function of a $\mathcal{N}_n(\mu, \boldsymbol{\Sigma})$ random vector with respect to the Lebesgue measure on $\mathbb{R}^n$, $\pi_0(\boldsymbol{\beta}_a, \phi)$ and $\pi_1(\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi)$ are the prior densities under the null and alternative hypotheses, respectively. If B_{10} is large, the alternative hypothesis is prefered. The behavior of a Bayes factor largely depends on the choice of priors. In Bayesian literature, many priors have been considered for testing the coefficients of linear model. Popular priors include g-priors (Liang et al., 2008) and intrinsic priors (Casella and Moreno, 2006). Unfortunately, these priors are not well defined in large p, small q setting. To obtain valid Bayes factor, we use simple priors.

Note that under the null hypothesis $\mathcal{H}_0$, the model is low-dimensional. This allows us to impose the reference prior $\pi_0(\boldsymbol{\beta}_a, \phi) = c/\phi$, where c is a constant. Under $\mathcal{H}_1$, write $\pi_1(\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi) = \pi_1(\boldsymbol{\beta}_b|\boldsymbol{\beta}_a, \phi)\pi_1(\boldsymbol{\beta}_a, \phi)$. For parameters $\boldsymbol{\beta}_a, \phi$, we consider the same prior as in $\mathcal{H}_0$, that is $\pi_1(\boldsymbol{\beta}_a, \phi) = \pi_0(\boldsymbol{\beta}_a, \phi)$. For parameter $\boldsymbol{\beta}_b$, however, imposing the improper reference prior will result in infinite marginal density of $\mathbf{y}$. To make the marginal density of $\mathbf{y}$ finite, we consider the simple normal prior $p_1(\boldsymbol{\beta}_b|\boldsymbol{\beta}_a, \phi) = d\mathcal{N}_p(0, \kappa^{-1}\phi^{-1}\mathbf{I}_p)(\boldsymbol{\beta}_b)$, where $\kappa > 0$ is a hyperparameter. That is, we put the following priors,

$$\pi_0(\boldsymbol{\beta}_a, \phi) = \frac{c}{\phi}, \quad \pi_1(\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi) = \frac{c}{\phi} d\mathcal{N}_p\left(0, \frac{1}{\kappa\phi}\mathbf{I}_p\right)(\boldsymbol{\beta}_b). \tag{3}$$

In what follows, we assume $\mathrm{Rank}(\mathbf{X}_a) = q$ and $\mathrm{Rank}([\mathbf{X}_a; \mathbf{X}_b]) = n$. Let $\mathbf{P}_a = \mathbf{X}_a(\mathbf{X}_a^\top\mathbf{X}_a)^{-1}\mathbf{X}^\top$ be the projection matrix onto the column space of $\mathbf{X}_a$. Let $B_{10,\kappa}$ be the Bayes factor corresponding to the priors (3). It is straightforward to show that

$$
\begin{aligned}
2\log(B_{10,\kappa}) =& p\log\kappa - \log|\mathbf{X}_b^\top(\mathbf{I}_n-\mathbf{P}_a)\mathbf{X}_b+\kappa\mathbf{I}_p| \\
&- (n-q)\log\left(1-\frac{\mathbf{y}\top(\mathbf{I}_n-\mathbf{P}_a)\mathbf{X}_b\left(\mathbf{X}_b^\top(\mathbf{I}-\mathbf{P}_a)\mathbf{X}_b+\kappa\mathbf{I}_p\right)^{-1}\mathbf{X}_b^\top(\mathbf{I}_n-\mathbf{P}_a)\mathbf{y}}{\mathbf{y}\top(\mathbf{I}_n-\mathbf{P}_a)\mathbf{y}}\right).
\end{aligned}
$$

Denote by $\mathbf{I}_n-\mathbf{P}_a=\tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top$ the rank decomposition of $\mathbf{I}_n-\mathbf{P}_a$, where $\tilde{\mathbf{U}}_a$ is a $n\times(n-q)$ column orthogonal matrix. Let $\mathbf{X}_b^*=\tilde{\mathbf{U}}_a^\top\mathbf{X}_b$, $\mathbf{y}^*=\tilde{\mathbf{U}}_a^\top\mathbf{y}$. Let γ_i be the ith largest eigenvalue of $\mathbf{X}_b^*\mathbf{X}_b^{*\top}$, $i=1,\ldots,n-q$. Denote by $\mathbf{X}_b^*=\mathbf{U}_b^*\mathbf{D}_b^*\mathbf{V}_b^{*\top}$ the singular value decomposition of $\mathbf{X}_b^*$, where $\mathbf{U}_b^*$, $\mathbf{V}_b^*$ are $(n-q)\times(n-q)$ and $p\times(n-q)$ column orthogonal matrices, respectively, and $\mathbf{D}_b^*=\operatorname{diag}(\sqrt{\gamma_1},\ldots,\sqrt{\gamma_{n-q}})$. Then we have

$$
2\log(B_{10,\kappa}) =(n-q)\log\kappa-\sum_{i=1}^{n-q}\log(\gamma_i+\kappa)-(n-q)\log\left(1-\frac{\mathbf{y}^{*\top}\mathbf{U}_b^*\mathbf{D}_b^*\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^*\mathbf{U}_b^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*}\right).
$$

The main part of the above expression is

$$
T_\kappa=\frac{\mathbf{y}^{*\top}\mathbf{U}_b^*\mathbf{D}_b^*\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^*\mathbf{U}_b^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*}.
$$

A large value of T_κ supports the alternative hypothesis. Hence T_κ can be regarded as a frequentist test statistic. The remaining problem is how to choose the hyperparameter κ. As κ increases, the prior magnitude of $\boldsymbol{\beta}_b$ decreases. As Goeman et al. (2006) noted, the priors should place most probability on the alternatives which are perceived as more interesting to detect. Their stretagy is to let the prior magnitude tend to zero to obtain a score test. In fact, if we let κ tends to infinity, the limit

$$
\lim_{\kappa\to\infty}\kappa T_\kappa=\frac{\mathbf{y}^{*\top}\mathbf{X}_b^*\mathbf{X}_b^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*}
$$

is exactly the test statistic of Goeman et al. (2006). This statistic has good power behavior under local alternatives, that is, $\|\boldsymbol{\beta}_b\|$ is small. As implied by Proposition 1, however, testing hypotheses (2) in large p, small q setting is a difficult problem. Intuitively, it may be too ambitious to talk about local power provided the test may have negligible power even when $\|\boldsymbol{\beta}_b\|$ is large.

Now we take further look at the statistics T_κ. Under the null hypothesis,

$$
\operatorname{E}T_\kappa=\frac{1}{n-q}\operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q})^{-1}\right).
$$

Under the alternative hypothesis, we consider the limiting behaivior of T_κ when $\|\boldsymbol{\beta}_b\|$ tends to infinity. Let $\boldsymbol{\beta}_b=c\boldsymbol{\beta}_b^\dagger$ where $\boldsymbol{\beta}_b^\dagger\neq 0$ is fixed and $c>0$. As $c\to\infty$,

$$
T_\kappa\to\frac{\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^*\mathbf{D}_b^{*2}\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*2}\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^\dagger}{\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^*\mathbf{D}_b^{*2}\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^\dagger}.
$$

Then the test $T_{n,\kappa}$ is unbiased along the direction of $\boldsymbol{\beta}_b^\dagger$ if

$$
\frac{\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^*\mathbf{D}_b^{*2}\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*2}\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^\dagger}{\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^*\mathbf{D}_b^{*2}\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^\dagger}>\frac{1}{n-q}\operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q})^{-1}\right),
$$

or equivalently

$$\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^*\left[\mathbf{D}_b^{*2}\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*2}-\frac{1}{n-q}\operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q})^{-1}\right)\mathbf{D}_b^{*2}\right]\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^{\dagger}>0. \tag{4}$$

Let k_κ be the number of positive eigenvalues of

$$\mathbf{V}_b^*\left[\mathbf{D}_b^{*2}\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*2}-\frac{1}{n-q}\operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q})^{-1}\right)\mathbf{D}_b^{*2}\right]\mathbf{V}_b^{*\top}.$$

Let $\mathcal{S}_\kappa$ be the linear space spanned by the first k_κ columns of $\mathbf{V}_b^*$. Denote by $\mathcal{S}_\kappa^\perp$ the orthogonal complement space of $\mathcal{S}_\kappa$. Then $\mathbb{R}^p=\mathcal{S}_\kappa\oplus\mathcal{S}_\kappa^\perp$. It can be seen that (4) holds for $\boldsymbol{\beta}_b^\dagger\in\mathcal{S}_\kappa$ and (4) does not hold for $\boldsymbol{\beta}_b^\dagger\in\mathcal{S}_\kappa^\perp$. Hence to achieve maximum unbiased directions, we would like to choose the κ which maximize k_κ.

Proposition 2. *The dimension k_κ of $\mathcal{S}_\kappa$ is nonincreasing in κ for $\kappa>0$. That is, $k_{\kappa_1}\geq k_{\kappa_2}$ for $\kappa_2>\kappa_1>0$.*

Proposition 2 implies that we should let κ tend to 0 to achieve maximum unbiased directions. While the statistic T_κ itself degenerates to 1 as $\kappa\to 0$, the right derivative of T_κ at $\kappa=0$ is well defined. Thus, we proposed the following test statistic

$$T=\left.\frac{\mathrm{d}T_{n,\kappa}}{\mathrm{d}\kappa}\right|_{\kappa=0}=-\frac{\mathbf{y}^{*\top}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*}.$$

The null hypothesis will be rejected if T is large.

3 Critical value

To formulate a valid frequentist test, we need to determine the critical value of T. If the data were indeed from normal distribution, then under the null hypohtesis,

$$T\sim-\frac{\sum_{i=1}^{n-q}\gamma_i^{-1}z_i^2}{\sum_{i=1}^{n-q}z_i^2},$$

where $z_1,\ldots,z_{n-q}$ are iid $\mathcal{N}(0,1)$ random variables. And the exact critical value can be easily obtained. However, normal distribution rarely appears in practice. We would like to derive an asymptotic valid critical value for T_n for general distributions of ϵ.

Under the null hypothesis,

$$T=-\frac{(\sqrt{\phi}\boldsymbol{\varepsilon})^\top\tilde{\mathbf{U}}_a(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\tilde{\mathbf{U}}_a^\top(\sqrt{\phi}\boldsymbol{\varepsilon})}{(\sqrt{\phi}\boldsymbol{\varepsilon})^\top\tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top(\sqrt{\phi}\boldsymbol{\varepsilon})}.$$

The numerator and the denominator of the above expression are both quadratic forms of iid random variables with mean 0 and variance 1. Hence our first step towards the goal is to approximate the distribution of the quadratic form of iid standardized random variables. The asymptotics of quaatratic form have been extensively studied; see, e.g., Jiang (1996); Bentkus and Gotze (1996); Götze and Tikhomirov (2002); Dicker and Erdogdu (2017); Bai et al. (2018). Most existing work focus on normal approximation. However, normal is just one of the possible limit distributions of quadratic form. See Sevast'yanov (1961) for a full characterazation of the limit distributions of quadratic form of normal random variables. Our approximation stretegy is to replace the random variables in quadratic form by suitable normal random variables. The error bound of this approximation will be derived by Lindeberg's replacement trick (see, e.g., Pollard (1984), Section III.4).

Let $\mathscr{C}^4(\mathbb{R})$ denote the class of all bounded real functions on $\mathbb{R}$ having bounded, continuous kth derivatives, $1 \leq k \leq 4$. It is known that if $\mathrm{E}\, f(Z_n) \to \mathrm{E}\, f(Z)$ for every $f \in \mathscr{C}^4(\mathbb{R})$ then $Z_n \rightsquigarrow Z$; see, e.g., Pollard (1984), Theorem 12 of Chapter III. We have the following approximation theorem.

Theorem 1. *Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^\top$, where ξ_i's are iid random variable with $\mathrm{E}\,\xi_1 = 0$, $\mathrm{Var}(\xi_1) = 1$. Furthermore, suppose the distribution of ξ_1 is symmetric about 0 and has finite eighth moments. Let $\mathbf{A}$ be an $n \times n$ symmetric matrix with elements $a_{i,j}$. Define*

$$S = \frac{\boldsymbol{\xi}^\top \mathbf{A} \boldsymbol{\xi} - \operatorname{tr}(\mathbf{A})}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\sum_{i=1}^n a_{i,i}^2}}.$$

Let $z_1, \ldots, z_n$ be iid random variables with distribution $\mathcal{N}(0,1)$ and $\check{z}_1, \ldots, \check{z}_n$ be iid random variables with distribution $\mathcal{N}(0,1)$ which are independent of $\xi_1, \ldots, \xi_n$. Let τ be a real number. Define

$$S_\tau^* = \frac{\tau \sum_{i=1}^n a_{i,i}\check{z}_i + 2\sum_{1 \leq i < j \leq n} a_{i,j} z_i z_j}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\sum_{i=1}^n a_{i,i}^2}}.$$

Then for every $f \in \mathscr{C}^4(\mathbb{R})$,

$$\begin{aligned}
&|\mathrm{E}\, f(S) - \mathrm{E}\, f(S_\tau^*)| \\
\leq &\frac{\left|\mathrm{E}(\xi_1^4) - 1 - \tau^2\right|}{2} \|f^{(2)}\|_\infty \frac{\sum_{i=1}^n a_{i,i}^2}{2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\sum_{i=1}^n a_{i,i}^2} \\
&+ \frac{\max\left(\left|\mathrm{E}(\xi_1^2 - 1)^3\right|, 12(\mathrm{E}(\xi_1^4) - 1)\right)}{6} \|f^{(3)}\|_\infty \frac{\sum_{l=1}^n \left(|a_{l,l}| \sum_{i=1}^n a_{i,l}^2\right)}{\left(2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\sum_{i=1}^n a_{i,i}^2\right)^{3/2}} \\
&+ \frac{16\,\mathrm{E}(\xi_1^8) + 80\,\mathrm{E}(\xi_1^4) + 3\tau^4 + 96}{24} \|f^{(4)}\|_\infty \frac{\sum_{l=1}^n \left(\sum_{i=1}^n a_{i,l}^2\right)^2}{\left(2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\sum_{i=1}^n a_{i,i}^2\right)^2}.
\end{aligned} \tag{5}$$

Remark 1. If $\tau^2 = \mathrm{E}(\xi_1^4) - 1$, the first term of the right hand side of (5) disappear. In practice, however, the quantity $\mathrm{E}(\xi_1^4)$ is often unknown and τ^2 should be chosen as an estimator of $\mathrm{E}(\xi_1^4) - 1$.

Remark 2. As noted in Chatterjee (2008), Section 3.1, an almost necessary condition for the asymptotic normality of S is

$$\frac{\operatorname{tr}(\mathbf{A}^4)}{\left(2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\sum_{i=1}^n a_{i,i}^2\right)^2} \to 0. \tag{6}$$

On the other hand, it can be seen that the right hand side of (5) converges to 0 provided $\tau^2 = \mathrm{E}(\xi_1^4) - 1$ and

$$\frac{\sum_{l=1}^n \left(\sum_{i=1}^n a_{i,l}^2\right)^2}{\left(2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\sum_{i=1}^n a_{i,i}^2\right)^2} \to 0. \tag{7}$$

It can be seen that (7) is much weaker than (6). For example, if $a_{i,j} = 1$, $i = 1, \ldots, n$, $j = 1, \ldots, n$ and $\mathrm{E}(\xi_1^4) = 3$, then the condition (7) holds but the condition (6) does not hold.

We now apply Theorem 1 to approximate the null distribution of the proposed statistic T. Note that under the null hypothesis,

$$T + \frac{\operatorname{tr}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\right)}{n-q} = \frac{(\sqrt{\phi}\boldsymbol{\varepsilon})^\top \left(-\tilde{\mathbf{U}}_a(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\tilde{\mathbf{U}}_a^\top + \frac{\operatorname{tr}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\right)}{n-q}\tilde{\mathbf{U}}_b\tilde{\mathbf{U}}_b^\top\right)(\sqrt{\phi}\boldsymbol{\varepsilon})}{(\sqrt{\phi}\boldsymbol{\varepsilon})^\top\tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top(\sqrt{\phi}\boldsymbol{\varepsilon})},$$

where the numerator has zero mean and the denominator is close to $n - q$. In what follows, let $\boldsymbol{\xi} = \sqrt{\phi}\boldsymbol{\varepsilon}$ and

$$\mathbf{A} = -\tilde{\mathbf{U}}_b(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\tilde{\mathbf{U}}_b^\top + \frac{\operatorname{tr}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\right)}{n-q}\tilde{\mathbf{U}}_b\tilde{\mathbf{U}}_b^\top.$$

Let $F_\tau(x)$ be the cumulative distribution function of $\tau\sum_{i=1}^n a_{i,i}\check{z}_i + 2\sum_{1\le i<j\le n} a_{i,j}z_iz_j$. As noted in Remark 1, τ^2 should be a consistent estimator of $\phi^2\,\mathrm{E}(\epsilon_1^4) - 1$ under the null hypothesis.

Theorem 2. *Suppose the conditions of Theorem 1 hold. Furthermore, suppose as $n \to \infty$, the condition (7) holds. Let $\hat{\tau}^2$ be an consistent estimator of $\phi^2\,\mathrm{E}(\epsilon_1^4) - 1$ based on $\mathbf{X}$, $\mathbf{y}$. Then*

$$\Pr\left(T > \frac{F_{\hat{\tau}}^{-1}(1-\alpha) - \operatorname{tr}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\right)}{n-q}\right) \to \alpha.$$

A consistent estimator of $\sigma^{-4}\,\mathrm{E}(\epsilon_1^4) - 1$ has already appeared in Bai et al. (2018) based on the standardized residuals. Here we use a slightly different estimator which is based on the ordinary least squares residuals $\tilde{\boldsymbol{\varepsilon}} = (\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_n)^\top = (\mathbf{I}_n - \mathbf{P}_a)\mathbf{y}$. From Bai et al. (2018), Theorem 2.1,

$$\begin{aligned}
&\mathrm{E}\left(\tilde{\boldsymbol{\varepsilon}}^\top(\mathbf{I}_n - \mathbf{P}_a)\tilde{\boldsymbol{\varepsilon}}\right) = (n-q)\sigma^2,\\
&\mathrm{E}\left(\sum_{i=1}^n \tilde{\epsilon}_i^4\right) = 3\sigma^4\operatorname{tr}(\mathbf{I}_n - \mathbf{P}_a)^{\circ 2} + \left(\mathrm{E}(\epsilon_1^4) - 3\sigma^4\right)\operatorname{tr}\left((\mathbf{I}_n - \mathbf{P}_a)^{\circ 2}\right)^2.
\end{aligned}$$

Then a moment estimator of $\sigma^{-4}\,\mathrm{E}(\epsilon_1^4) - 1$ is

$$\hat{\tau}^2 = \frac{\dfrac{(n-q)^2\sum_{i=1}^n \tilde{\epsilon}_i^4}{\left(\tilde{\boldsymbol{\varepsilon}}^\top(\mathbf{I}_n - \mathbf{P}_a)\tilde{\boldsymbol{\varepsilon}}\right)^2} - 3\operatorname{tr}(\mathbf{I}_n - \mathbf{P}_a)^{\circ 2}}{\operatorname{tr}\left((\mathbf{I}_n - \mathbf{P}_a)^{\circ 2}\right)^2} + 2.$$

Proposition 3. *Suppose the conditions of 1 holds for $\boldsymbol{\xi} = \sqrt{\phi}\boldsymbol{\varepsilon}$. Suppose $q/n \to 0$. Then under the null hypothesis, $\hat{\tau}^2 \xrightarrow{P} \sigma^{-4} \operatorname{E}(\epsilon_1^4) - 1$.*

We reject the null hypothesis if

$$T > \frac{F_{\hat{\tau}}^{-1}(1-\alpha) - \operatorname{tr}\left((\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1}\right)}{n-q}.$$

This test procedure is asymptotically exact of size α under the conditions of Theorem 2 and Proposition 3.

4 Power analysis

In this section, we investigate the power behavior of the proposed test procedure under a special case.

Let

$$R_k = \frac{\mathbf{y}^{*\top}(\mathbf{X}_b^* \mathbf{X}_b^{*\top})^k \mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*} - \frac{\operatorname{tr}(\mathbf{X}_b^* \mathbf{X}_b^{*\top})^k}{n-q}.$$

Theorem 3. *Suppose the rows of $\mathbf{X}_b$ are iid random vectors with distribution $\mathcal{N}_p(0, \sigma_b^2 \mathbf{I}_p)$. Suppose $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n)$. Suppose $k \neq 0$. Suppose $n - q \to \infty$, $p/(n-q) \to c \in (1, +\infty)$. Then*

$$\sup_{x\in\mathbb{R}} \left| \Pr\left(R_k \leq x\right) - \Phi\left(\frac{\sqrt{n-q}}{\sqrt{2\sigma_b^{4k} p^{2k} \operatorname{E}(\eta^k - \operatorname{E}\eta^k)^2}} \left(x - \|\boldsymbol{\beta}_b\|^2 \phi \sigma_b^{2k+2} p^k \left(\operatorname{E}\eta^{k+1} - \operatorname{E}\eta^k \operatorname{E}\eta \right)\right)\right)\right| \to 0.$$

The proposed test reject the null hypothesis if R_{-1} is small. Their test reject the null hypothesis if R_1 is large.

Under the null hypothesis, we have

$$\Pr\left(R_{-1} \leq \frac{\sqrt{2\sigma_b^{-4} p^{-2} \operatorname{E}(\eta^{-1} - \operatorname{E}\eta^{-1})^2}}{\sqrt{n-q}} \Phi^{-1}(\alpha) \right) \to \alpha.$$

Under the alternative hypothesis,

$$\Pr\left(R_{-1} \leq \frac{\sqrt{2\sigma_b^{-4} p^{-2} \operatorname{E}(\eta^{-1} - \operatorname{E}\eta^{-1})^2}}{\sqrt{n-q}} \Phi^{-1}(\alpha) \right)$$

And

$$\Pr\left(R_1 > \frac{\sqrt{2\sigma_b^{4} p^{2} \operatorname{E}(\eta - \operatorname{E}\eta)^2}}{\sqrt{n-q}} \Phi^{-1}(1-\alpha) \right) \to \alpha.$$

Then the power function is

$$
\begin{aligned}
&\Pr\left(T_n - \operatorname{tr}((\mathbf{X}_b^* \mathbf{X}_b^{*\top})^k)/(n-q) \le \Phi(\alpha)\frac{\sqrt{2\sigma_b^{-4}p^{-2}(n-q)(\nu_{-2,c}-\nu_{-1,c}^2)}}{n-q}\right)\\
=&\Pr\left(\mathcal{N}(s,1) \le \Phi(\alpha)\frac{(n-q)\left(1+\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2\nu_{1,c}\right)\sqrt{2\sigma_b^{-4}p^{-2}(n-q)(\nu_{-2,c}-\nu_{-1,c}^2)}}{(n-q)\sqrt{2\sigma_b^{-2}p^{-2}(n-q)\left(\sigma_b^{-2}(\nu_{-2,c}-\nu_{-1,c}^2)+2\phi\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2-\nu_{-1,c})\right)}}\right)+o(1)\\
=&\Pr\left(\mathcal{N}(s,1) \le \Phi(\alpha)\frac{\left(1+\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2\nu_{1,c}\right)}{\sqrt{1+\frac{2\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2-\nu_{-1,c})}{\nu_{-2,c}-\nu_{-1,c}^2}}}\right)+o(1)\\
=&\Phi\left(-s+\Phi(\alpha)\frac{\left(1+\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2\nu_{1,c}\right)}{\sqrt{1+\frac{2\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2-\nu_{-1,c})}{\nu_{-2,c}-\nu_{-1,c}^2}}}\right)+o(1)\\
=&\Phi\left(\frac{\phi\|\boldsymbol{\beta}_b\|^2p^{-1}(n-q)\left(\nu_{1,c}\nu_{-1,c}-1\right)}{\sqrt{2\sigma_b^{-2}p^{-2}(n-q)\left(\sigma_b^{-2}(\nu_{-2,c}-\nu_{-1,c}^2)+2\phi\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2-\nu_{-1,c})\right)}}+\Phi(\alpha)\frac{\left(1+\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2\nu_{1,c}\right)}{\sqrt{1+\frac{2\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2-\nu_{-1,c})}{\nu_{-2,c}-\nu_{-1,c}^2}}}\right)\\
=&\Phi\left(\frac{\sqrt{n-q}\,\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2\left(\nu_{1,c}\nu_{-1,c}-1\right)}{\sqrt{2\left((\nu_{-2,c}-\nu_{-1,c}^2)+2\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2-\nu_{-1,c})\right)}}+\Phi(\alpha)\frac{\left(1+\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2\nu_{1,c}\right)}{\sqrt{1+\frac{2\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2-\nu_{-1,c})}{\nu_{-2,c}-\nu_{-1,c}^2}}}\right)+o(1)
\end{aligned}
$$

Appendices

Appendix A Proofs of the results in Section 2

Proof of Proposition 1. Since $\operatorname{Rank}([\mathbf{X}_a, \mathbf{X}_b]) = n$, $\phi(\mathbf{y})$ is unbiased if and only if

$$\int_{\mathbb{R}^n} \varphi(\mathbf{y}) d\mathcal{N}_n(\mu, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\, \mathrm{d}\mathbf{y} \ge \alpha \quad \text{for all } \mu \in \mathbb{R}^n.$$

From E. L. Lehmann (2005), Theorem 2.7.1, $\mathrm{E}[\phi(\mathbf{y})] = \alpha$ under the null hypothesis. In particular, we have

$$\int_{\mathbb{R}^n} [\varphi(\mathbf{y}) - \alpha] d\mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\, \mathrm{d}\mathbf{y} = 0. \tag{8}$$

If $\alpha = 0$ or 1, the conclusion is trivially true. In what follows, we assume $0 < \alpha < 1$. We claim that if $\varphi(\mathbf{y}) \ge \alpha$, a.s. λ, then the conclusion holds. In fact, if $\varphi(\mathbf{y}) \ge \alpha$, a.s. λ, then the integrand of (8) is nonnegative, and hence must be 0 a.s. λ, which implies the conclusion. Next we prove $\varphi(\mathbf{y}) \ge \alpha$, a.s. λ by contradiction. Suppose $\lambda(\{\mathbf{y} : \varphi(\mathbf{y}) < \alpha\}) > 0$. Then there exists a sufficiently small $\eta > 0$, such that $\lambda(\{\mathbf{y} : \varphi(\mathbf{y}) < \alpha - \eta\}) > 0$. We denote $E := \{\mathbf{y} \in \mathbb{R}^n : \varphi(\mathbf{y}) < \alpha - \eta\}$. From

Lebesgue density theorem (Cohn, 2013, Corollary 6.2.6), there exists a point $z \in E$, such that, for any $\epsilon > 0$ there is a $\delta_\epsilon > 0$ such that

$$\left| \frac{\lambda(E^{\complement} \cap C_\epsilon)}{\lambda(C_\epsilon)} \right| < \epsilon,$$

where $C_\epsilon = \prod_{i=1}^n [z_i - \delta_\epsilon, z_i + \delta_\epsilon]$. We put

$$\epsilon = \left(\frac{\sqrt{\pi}}{\sqrt{2}\Phi^{-1}\left(1 - \frac{\eta}{6n}\right)} \right)^n \frac{\eta}{3},$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable. Then for any $\phi > 0$,

$$\begin{aligned}
\alpha \leq & \int_{\mathbb{R}^n} \varphi(\mathbf{y}) d\mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(\mathbf{y}) \, \mathrm{d}\mathbf{y} \\
= & \int_{E \cap C_\epsilon} \varphi(\mathbf{y}) d\mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(\mathbf{y}) \, \mathrm{d}\mathbf{y} + \int_{E^{\complement} \cap C_\epsilon} \varphi(\mathbf{y}) d\mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(\mathbf{y}) \, \mathrm{d}\mathbf{y} + \int_{C_\epsilon^{\complement}} \varphi(\mathbf{y}) d\mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(\mathbf{y}) \, \mathrm{d}\mathbf{y} \\
\leq & \alpha - \eta + \int_{E^{\complement} \cap C_\epsilon} d\mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(\mathbf{y}) \, \mathrm{d}\mathbf{y} + \int_{C_\epsilon^{\complement}} d\mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(\mathbf{y}) \, \mathrm{d}\mathbf{y} \\
\leq & \alpha - \eta + \left(\frac{\phi}{2\pi} \right)^{n/2} \lambda(E^{\complement} \cap C_\epsilon) + 2n \left(1 - \Phi(\sqrt{\phi}\delta_\epsilon) \right) \\
\leq & \alpha - \eta + \left(\frac{\phi}{2\pi} \right)^{n/2} \epsilon (2\delta_\epsilon)^n + 2n \left(1 - \Phi(\sqrt{\phi}\delta_\epsilon) \right) \\
= & \alpha - \eta + \left(\frac{\sqrt{\phi}\delta_\epsilon}{\Phi^{-1}\left(1 - \frac{\eta}{6n}\right)} \right)^n \frac{\eta}{3} + 2n \left(1 - \Phi(\sqrt{\phi}\delta_\epsilon) \right).
\end{aligned}$$

Putting

$$\phi = \left(\frac{\Phi^{-1}\left(1 - \frac{\eta}{6n}\right)}{\delta_\epsilon} \right)^2$$

yields the contradiction $\alpha \leq \alpha - (2/3)\eta$. This completes the proof.

$\square$

Proof of Proposition 2. For positive integer m, define $[m] = \{1, , \ldots, m\}$. For a set A, denote by $|A|$ its cardinality. We have

$$\begin{aligned}
k_\kappa = & \left| \left\{ i \in [n-q] : \frac{\gamma_i^2}{\gamma_i + \kappa} - \frac{1}{n-q} \sum_{j=1}^{n-q} \frac{\gamma_j \gamma_i}{\gamma_j + \kappa} > 0 \right\} \right| \\
= & \left| \left\{ i \in [n-q] : \frac{\gamma_i}{\gamma_i + \kappa} > \frac{1}{n-q} \sum_{j=1}^{n-q} \frac{\gamma_j}{\gamma_j + \kappa} \right\} \right|.
\end{aligned}$$

Let X be a random variable uniformly distributed on $\{\gamma_1, \ldots, \gamma_{n-q}\}$. That is, $\Pr(X = \gamma_i) = 1/(n-q)$, $i = 1, \ldots, n-q$. Then it can be seen that

$$k_\kappa = (n-q) \Pr\left(\frac{X}{X + \kappa} > \mathrm{E}\left[\frac{X}{X + \kappa} \right] \right).$$

Hence we only need to verify

$$\Pr\left(\frac{X}{X+\kappa_1} > \mathrm{E}\left[\frac{X}{X+\kappa_1}\right]\right) \geq \Pr\left(\frac{X}{X+\kappa_2} > \mathrm{E}\left[\frac{X}{X+\kappa_2}\right]\right). \tag{9}$$

Let $Y = X/(X+\kappa_2)$. Then

$$\frac{X}{(X+\kappa_1)} = \frac{\kappa_2 Y}{\kappa_1 + (\kappa_2 - \kappa_1)Y} =: f(Y).$$

Note that $f(Y)$ is increasing for $Y \geq 0$. Then the inequality (9) is equivalent to

$$\Pr\left(Y > f^{-1}\left(\mathrm{E}\, f(Y)\right)\right) \geq \Pr\left(Y > \mathrm{E}\, Y\right).$$

Hence we only need to verify $f^{-1}\left(\mathrm{E}\, f(Y)\right) \leq \mathrm{E}\, Y$, or equivalently, $\mathrm{E}\, f(Y) \leq f(\mathrm{E}\, Y)$. But the last inequality is a direct consequence of the concavity of $f(Y)$ and Jensen's inequality. This completes the proof.

□

Appendix B haha3

Proof of Theorem 1. Let

$$\tilde{a}_{i,j} := \frac{a_{i,j}}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\operatorname{tr}(\mathbf{A} \circ \mathbf{A})}}.$$

Then

$$S = \sum_{i=1}^{n} \tilde{a}_{i,i}(\xi_i^2 - 1) + 2 \sum_{1 \leq i < j \leq n} \tilde{a}_{i,j}\xi_i\xi_j, \quad S_\tau^* = \tau \sum_{i=1}^{n} \tilde{a}_{i,i}\check{z}_i + 2 \sum_{1 \leq i < j \leq n} \tilde{a}_{i,j} z_i z_j.$$

For $l = 1, \ldots, n$, define

$$\begin{aligned}
S_l =& \sum_{i=1}^{l-1} \tilde{a}_{i,i}(\xi_i^2 - 1) + \tau \sum_{i=l+1}^{n} \tilde{a}_{i,i}\check{z}_i + 2 \sum_{1 \leq i < j \leq l-1} \tilde{a}_{i,j}\xi_i\xi_j + 2 \sum_{i=1}^{l-1} \sum_{j=l+1}^{n} \tilde{a}_{i,j}\xi_i z_j + 2 \sum_{l+1 \leq i < j \leq n} \tilde{a}_{i,j} z_i z_j, \\
h_l =& \tilde{a}_{l,l}(\xi_l^2 - 1) + 2 \sum_{i=1}^{l-1} \tilde{a}_{i,l}\xi_i\xi_l + 2 \sum_{i=l+1}^{n} \tilde{a}_{i,l} z_i \xi_l, \\
g_l =& \tau\tilde{a}_{l,l}\check{z}_l + 2 \sum_{i=1}^{l-1} \tilde{a}_{i,l}\xi_i z_l + 2 \sum_{i=l+1}^{n} \tilde{a}_{i,l} z_i z_l.
\end{aligned}$$

It can be seen that for $l = 2, \ldots, n$, $S_{l-1} + h_{l-1} = S_l + g_l$, and $S = S_n + h_n$, $S_1 + g_1 = S_\tau^*$.

Thus, for any $f \in \mathscr{C}^4(\mathbb{R})$,

$$\begin{aligned}
|\mathrm{E}\, f(S) - \mathrm{E}\, f(S_\tau^*)| &= |\mathrm{E}\, f(S_n + h_n) - \mathrm{E}\, f(S_1 + g_1)| \\
&= \left| \sum_{l=2}^{n} \left(\mathrm{E}\, f(S_l + h_l) - \mathrm{E}\, f(S_{l-1} + h_{l-1})\right) + \mathrm{E}\, f(S_1 + h_1) - \mathrm{E}\, f(S_1 + g_1) \right| \\
&= \left| \sum_{l=1}^{n} \mathrm{E}\, f(S_l + h_l) - \mathrm{E}\, f(S_l + g_l) \right|.
\end{aligned}$$

Apply Taylor's theorem, for $l = 1, \dots, n$,

$$\begin{aligned}
f(S_l + h_l) =& f(S_l) + \sum_{k=1}^{3} \frac{1}{k!} h_l^k f^{(k)}(S_l) + \frac{1}{24} h_l^4 f^{(4)}(S_l + \theta_1 h_l), \\
f(S_l + g_l) =& f(S_l) + \sum_{k=1}^{3} \frac{1}{k!} g_l^k f^{(k)}(S_l) + \frac{1}{24} g_l^4 f^{(4)}(S_l + \theta_2 g_l),
\end{aligned}$$

where $\theta_1, \theta_2 \in [0,1]$. Thus,

$$|\mathrm{E}\, f(S_l + h_l) - \mathrm{E}\, f(S_l + g_l)| \le \left| \sum_{k=1}^{3} \frac{1}{k!} \mathrm{E}\, f^{(k)}(S_l) \, \mathrm{E}_l(h_l^k - g_l^k) \right| + \frac{1}{24} \|f''''\|_\infty \left(\mathrm{E}(h_l^4) + \mathrm{E}(g_l^4) \right),$$

where E_l denotes taking expectation with respect to $\xi_l, z_l, \check{z}_l$. It is straightforward to show that

$$\begin{aligned}
&\mathrm{E}_l(h_l - g_l) = 0, \\
&\mathrm{E}_l(h_l^2 - g_l^2) = \left(\mathrm{E}(\xi_1^4) - 1 - \tau^2 \right) \tilde{a}_{l,l}^2, \\
&\mathrm{E}_l(h_l^3 - g_l^3) = \mathrm{E}(\xi_1^2 - 1)^3 \tilde{a}_{l,l}^3 + 12(\mathrm{E}(\xi_1^4) - 1) \tilde{a}_{l,l} \left(\sum_{i=1}^{l-1} \tilde{a}_{i,l} \xi_i + \sum_{i=l+1}^{n} \tilde{a}_{i,l} z_i \right)^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
&|\mathrm{E}\, f(S_l + h_l) - \mathrm{E}\, f(S_l + g_l)| \\
\le& \frac{1}{2} \|f^{(2)}\|_\infty \left| \mathrm{E}(\xi_1^4) - 1 - \tau^2 \right| \tilde{a}_{l,l}^2 \\
&+ \frac{1}{6} \|f^{(3)}\|_\infty \left(\left| \mathrm{E}(\xi_1^2 - 1)^3 \right| |\tilde{a}_{l,l}^3| + 12(\mathrm{E}(\xi_1^4) - 1) |\tilde{a}_{l,l}| \, \mathrm{E} \left(\sum_{i=1}^{l-1} \tilde{a}_{i,l} \xi_i + \sum_{i=l+1}^{n} \tilde{a}_{i,l} z_i \right)^2 \right) \\
&+ \frac{1}{24} \|f^{(4)}\|_\infty \left(\mathrm{E}(h_l^4) + \mathrm{E}(g_l^4) \right) \\
\le& \frac{\left| \mathrm{E}(\xi_1^4) - 1 - \tau^2 \right|}{2} \|f^{(2)}\|_\infty \tilde{a}_{l,l}^2 + \frac{\max\left(\left| \mathrm{E}(\xi_1^2 - 1)^3 \right|, 12(\mathrm{E}(\xi_1^4) - 1) \right)}{6} \|f^{(3)}\|_\infty |\tilde{a}_{l,l}| \sum_{i=1}^{n} \tilde{a}_{i,l}^2 \\
&+ \frac{1}{24} \|f^{(4)}\|_\infty \left(\mathrm{E}(h_l^4) + \mathrm{E}(g_l^4) \right).
\end{aligned} \tag{10}$$

Now we bound $\mathrm{E}(h_l^4)$ and $\mathrm{E}(g_l^4)$. By direct calculation,

$$\begin{aligned}
\mathrm{E}(h_l^4) =& \mathrm{E}(\xi_1^2 - 1)^4 \tilde{a}_{l,l}^4 + 24\, \mathrm{E}[\xi_1^2 (\xi_1^2 - 1)^2] \tilde{a}_{l,l}^2 \, \mathrm{E} \left(\sum_{i=1}^{l-1} \tilde{a}_{i,l} \xi_i + \sum_{i=l+1}^{n} \tilde{a}_{i,l} z_i \right)^2 \\
&+ 16\, \mathrm{E}(\xi_1^4) \, \mathrm{E} \left(\sum_{i=1}^{l-1} \tilde{a}_{i,l} \xi_i + \sum_{i=l+1}^{n} \tilde{a}_{i,l} z_i \right)^4 \\
=& \mathrm{E}(\xi_1^2 - 1)^4 \tilde{a}_{l,l}^4 + 24\, \mathrm{E}[\xi_1^2 (\xi_1^2 - 1)^2] \tilde{a}_{l,l}^2 \left((\sum_{i=1}^{n} \tilde{a}_{i,l}^2) - \tilde{a}_{l,l}^2 \right) \\
&+ 16\, \mathrm{E}(\xi_1^4) \left(\left(\mathrm{E}(\xi_1^4) - 3 \right) \sum_{i=1}^{l-1} \tilde{a}_{i,l}^4 + 3 \left((\sum_{i=1}^{n} \tilde{a}_{i,l}^2) - \tilde{a}_{l,l}^2 \right)^2 \right).
\end{aligned}$$

To upper bound the above quantity, we use the facts $24\,\mathrm{E}[\xi_1^2(\xi_1^2-1)^2] \le 2(16\,\mathrm{E}(\xi_1^2-1)^4+(9/4)\,\mathrm{E}(\xi_1^4))$, $\mathrm{E}(\xi_1^2-1)^4 \le \mathrm{E}(\xi_1^8)$ and

$$\big(\mathrm{E}(\xi_1^4)-3\big)\sum_{i=1}^{l-1}\tilde{a}_{i,l}^4 \le \big(\mathrm{E}(\xi_1^4)-1\big)\sum_{i=1}^{l-1}\tilde{a}_{i,l}^4 \le \big(\mathrm{E}(\xi_1^4)-1\big)\left((\sum_{i=1}^{n}\tilde{a}_{i,l}^2)-\tilde{a}_{l,l}^2\right)^2.$$

Then we obtain the bound

$$\mathrm{E}(h_l^4) \le \big(16\,\mathrm{E}(\xi_1^8)+32\,\mathrm{E}(\xi_1^4)\big)\left(\sum_{i=1}^{n}\tilde{a}_{i,l}^2\right)^2. \tag{11}$$

Similarly, we have

$$\mathrm{E}(g_l^4) \le \big(48\,\mathrm{E}(\xi_1^4)+3\tau^4+96\big)\left(\sum_{i=1}^{n}\tilde{a}_{i,l}^2\right)^2. \tag{12}$$

Combining (10), (11) and (12) yields

$$\begin{aligned}
&\sum_{l=1}^{n}\left|\mathrm{E}\,f(S_l+h_l)-\mathrm{E}\,f(S_l+g_l)\right| \\
\le&\frac{\left|\mathrm{E}(\xi_1^4)-1-\tau^2\right|}{2}\|f^{(2)}\|_\infty\sum_{l=1}^{n}\tilde{a}_{l,l}^2+\frac{\max\left(\left|\mathrm{E}(\xi_1^2-1)^3\right|,12(\mathrm{E}(\xi_1^4)-1)\right)}{6}\|f^{(3)}\|_\infty\sum_{l=1}^{n}\left(|\tilde{a}_{l,l}|\sum_{i=1}^{n}\tilde{a}_{i,l}^2\right) \\
&+\frac{16\,\mathrm{E}(\xi_1^8)+80\,\mathrm{E}(\xi_1^4)+3\tau^4+96}{24}\|f^{(4)}\|_\infty\sum_{l=1}^{n}\left(\sum_{i=1}^{n}\tilde{a}_{i,l}^2\right)^2.
\end{aligned}$$

This completes the proof.

$\square$

Proof of Theorem 2. By a standard subsequence argument, we only need to prove the theorem holds along a subsequence of $\{n\}$. Hence, without loss of generality, we assume $\hat{\tau}^2 \xrightarrow{a.s.} \phi^2\,\mathrm{E}(\epsilon_1^4)-1$. Note that almost surely, the distributions $\mathcal{L}(S_{\hat{\tau}}^*|\hat{\tau})$ have bounded second moment and are thus tight. Hence, without loss of generality, we assume $\mathcal{L}(S_{\hat{\tau}}^*|\hat{\tau})$ weakly converges to a limit distribution with distribution function $F(x)$. Let $S^\dagger$ be a random variable with this limit distribution. Similar to Chen et al. (2010), Proposition A.1.(iii), it can be shown that $\mathrm{E}[(S_{\hat{\tau}}^*)^4|\hat{\tau}]$ is uniformly bounded almost surely. Then almost surely, $\mathcal{L}((S_{\hat{\tau}}^*)^2|\hat{\tau})$ is uniformly integrable. Hence $F(x)$ can not concentrate on a single point. Consequently, $F(x)$ is continous and is strict increasing for $x \in \{x : 0 < F(x) < 1\}$; see Sevast'yanov (1961) as well as the remark made by A. N. Kolmogorov.

For every $f \in \mathscr{C}^4(\mathbb{R})$, we have $|\,\mathrm{E}[f(S_{\hat{\tau}}^*)|\hat{\tau}]-\mathrm{E}\,f(S^\dagger)| \to 0$ almost surely. Also, Theorem 1 implies $|\,\mathrm{E}\,f(S)-\mathrm{E}[f(S_{\hat{\tau}}^*)|\hat{\tau}]| \to 0$ almost surely. Thus, $|\,\mathrm{E}\,f(S)-\mathrm{E}\,f(S^\dagger)| \to 0$. That is, $S \rightsquigarrow S^\dagger$. On the other hand, since

$$\Pr\left(S_{\hat{\tau}}^* > \frac{F_{\hat{\tau}}^{-1}(1-\alpha)}{\sqrt{2\,\mathrm{tr}(\mathbf{A}^2)+(\phi^2\,\mathrm{E}(\epsilon_1^4)-3)\sum_{i=1}^{n}a_{i,i}^2}}\,\middle|\,\hat{\tau}\right)=\alpha,$$

we have almost surely,

$$\frac{F_{\hat{\tau}}^{-1}(1-\alpha)}{\sqrt{2\operatorname{tr}(\mathbf{A}^2)+(\phi^2\operatorname{E}(\epsilon_1^4)-3)\sum_{i=1}^n a_{i,i}^2}} \to F^{-1}(1-\alpha). \tag{13}$$

We also need the fact that

$$(\sqrt{\phi}\boldsymbol{\varepsilon})^\top \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\varepsilon}) = (1+o_p(1))(n-q), \tag{14}$$

which is a consequence of

$$\operatorname{E}\left((\sqrt{\phi}\boldsymbol{\varepsilon})^\top \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\varepsilon})\right) = n-q, \quad \operatorname{Var}\left((\sqrt{\phi}\boldsymbol{\varepsilon})^\top \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\varepsilon})\right) = O(n-q).$$

The fact $S \rightsquigarrow S^\dagger$, the equations (13), (14) and Slutsky's theorem leads to

$$\begin{aligned}
&\Pr\left(T > \frac{F_{\hat{\tau}}^{-1}(1-\alpha) - \operatorname{tr}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\right)}{n-q}\right) \\
=&\Pr\Bigg((\sqrt{\phi}\boldsymbol{\varepsilon})^\top\left(-\tilde{\mathbf{U}}_a(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\tilde{\mathbf{U}}_a^\top + \frac{\operatorname{tr}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\right)}{n-q}\tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top\right)(\sqrt{\phi}\boldsymbol{\varepsilon}) \\
&\qquad > \frac{(\sqrt{\phi}\boldsymbol{\varepsilon})^\top \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\varepsilon})}{n-q} F_{\hat{\tau}}^{-1}(1-\alpha)\Bigg) \\
=&\Pr\left(\boldsymbol{\xi}^\top\mathbf{A}\boldsymbol{\xi} > \frac{(\sqrt{\phi}\boldsymbol{\varepsilon})^\top \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\varepsilon})}{n-q} F_{\hat{\tau}}^{-1}(1-\alpha)\right) \\
=&\Pr\left(S > \frac{(\sqrt{\phi}\boldsymbol{\varepsilon})^\top \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\varepsilon})}{n-q} \frac{F_{\hat{\tau}}^{-1}(1-\alpha)}{\sqrt{2\operatorname{tr}(\mathbf{A}^2)+(\phi^2\operatorname{E}(\epsilon_1^4)-3)\sum_{i=1}^n a_{i,i}^2}}\right) \\
=&\Pr\Bigg(S + F^{-1}(1-\alpha) - \frac{(\sqrt{\phi}\boldsymbol{\varepsilon})^\top \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\varepsilon})}{n-q} \frac{F_{\hat{\tau}}^{-1}(1-\alpha)}{\sqrt{2\operatorname{tr}(\mathbf{A}^2)+(\phi^2\operatorname{E}(\epsilon_1^4)-3)\sum_{i=1}^n a_{i,i}^2}} \\
&\qquad > F^{-1}(1-\alpha)\Bigg) \\
\to&\alpha.
\end{aligned}$$

This completes the proof.

□

Proof of Proposition 3. From Bai et al. (2018), Theorem 2.1, one can obtain the explicit forms of $\operatorname{Var}\left(\tilde{\boldsymbol{\varepsilon}}^\top(\mathbf{I}_n-\mathbf{P}_a)\tilde{\boldsymbol{\varepsilon}}\right)$ and $\operatorname{Var}\left(\sum_{i=1}^n \tilde{\epsilon}_i^4\right)$ which involves the traces of certain matrices. Using Horn and Johnson (1991), Theorem 5.5.1, one can see that the eigenvalues of these matrices are all bounded. Hence it can be deduced that $\operatorname{Var}\left(\tilde{\boldsymbol{\varepsilon}}^\top(\mathbf{I}_n-\mathbf{P}_a)\tilde{\boldsymbol{\varepsilon}}\right) = O(n)$ and $\operatorname{Var}\left(\sum_{i=1}^n \tilde{\epsilon}_i^4\right) = O(n)$. Thus,

$$\begin{aligned}
&\tilde{\boldsymbol{\varepsilon}}^\top(\mathbf{I}_n-\mathbf{P}_a)\tilde{\boldsymbol{\varepsilon}} = (n-q)\sigma^2 + O_P(\sqrt{n}), \\
&\sum_{i=1}^n \tilde{\epsilon}_i^4 = 3\sigma^4\operatorname{tr}(\mathbf{I}_n-\mathbf{P}_a)^{\circ 2} + \left(\operatorname{E}(\epsilon_1^4)-3\sigma^4\right)\operatorname{tr}\left((\mathbf{I}_n-\mathbf{P}_a)^{\circ 2}\right)^2 + O_P(\sqrt{n}).
\end{aligned}$$

It follows that

$$\hat{\tau}^2 = \sigma^{-4}\operatorname{E}(\epsilon_1^4) - 1 + O_P\left(\frac{\sqrt{n}}{\operatorname{tr}\left((\mathbf{I}_n - \mathbf{P}_a)^{\circ 2}\right)^2}\right).$$

Let $\delta_{i,j} = 1$ if $i = j$ and 0 if $i \neq j$. We have

$$\begin{aligned}
n =& \sum_{i=1}^n \sum_{j=1}^n \delta_{i,j}^4 \\
=& \sum_{i=1}^n \sum_{j=1}^n (\delta_{i,j} - (\mathbf{P}_a)_{i,j} + (\mathbf{P}_a)_{i,j})^4 \\
\leq& 8 \sum_{i=1}^n \sum_{j=1}^n (\delta_{i,j} - (\mathbf{P}_a)_{i,j})^4 + 8 \sum_{i=1}^n \sum_{j=1}^n (\mathbf{P}_a)_{i,j}^4 \\
\leq& 8 \sum_{i=1}^n \sum_{j=1}^n (\delta_{i,j} - (\mathbf{P}_a)_{i,j})^4 + 8 \sum_{i=1}^n \sum_{j=1}^n (\mathbf{P}_a)_{i,j}^2 \\
=& 8\operatorname{tr}\left((\mathbf{I}_n - \mathbf{P}_a)^{\circ 2}\right)^2 + 8q.
\end{aligned}$$

Then

$$\frac{\sqrt{n}}{\operatorname{tr}\left((\mathbf{I}_n - \mathbf{P}_a)^{\circ 2}\right)^2} = O\left(\frac{1}{\sqrt{n}}\right).$$

This completes the proof. $\square$

Appendix C hahayy

Lemma 1. *Let $Z = (z_1, \ldots, z_n)^\top$, where z_i's are iid $\mathcal{N}(0,1)$ random variables. Let $\mathbf{A}$ be an $n \times n$ symmetric matrix with elements $a_{i,j}$. Let $\mathbf{b} = (b_1, \ldots, b_n)^\top$ be an n dimensional vector. If $\operatorname{tr}(\mathbf{A}^4)/\operatorname{tr}^2(\mathbf{A}^2) \to 0$, then*

$$\frac{Z^\top \mathbf{A} Z + b^\top Z - \operatorname{tr}(\mathbf{A})}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + \|\mathbf{b}\|^2}} \rightsquigarrow \mathcal{N}(0,1).$$

Proof. Without loss of generality, we assume $\mathbf{A}$ is a diagonal matrix and $|b_1| \geq \cdots \geq |b_n|$. By a standard subsequence argument, we only need to prove the result along a subsequence. Hence we can assume $\lim_{n\to\infty} \|b\|^2/\operatorname{tr}(\mathbf{A}^2) = c \in [0, +\infty]$. If $c = 0$, Lyapunov central limit theorem implies that

$$\frac{Z^\top \mathbf{A} Z + b^\top Z - \operatorname{tr}(\mathbf{A})}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + \|\mathbf{b}\|^2}} = (1 + o_P(1))\frac{Z^\top \mathbf{A} Z - \operatorname{tr}(\mathbf{A})}{\sqrt{2\operatorname{tr}(\mathbf{A}^2)}} + o_P(1) \rightsquigarrow \mathcal{N}(0,1).$$

If $c = +\infty$,

$$\frac{Z^\top \mathbf{A} Z + b^\top Z - \operatorname{tr}(\mathbf{A})}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + \|\mathbf{b}\|^2}} = (1 + o_P(1))\frac{b^\top Z}{\|\mathbf{b}\|} + o_P(1) \rightsquigarrow \mathcal{N}(0,1).$$

In what follows, we assume $c \in (0, +\infty)$. By Helly selection theorem, we can assume $\lim_{n\to\infty} |b_i|/\|\mathbf{b}\| = b_i^* \in [0,1]$, $i = 1, 2, \ldots$. From Fatou's lemma, we have $\sum_{i=1}^\infty (b_i^*)^2 \leq 1$. Consequently, $\lim_{i\to\infty} b_i^* = 0$.

Note that the condition $\operatorname{tr}(\mathbf{A}^4)/\operatorname{tr}^2(\mathbf{A}^2) \to 0$ is equivalent to $\lambda_1(\mathbf{A}^2)/\operatorname{tr}(\mathbf{A}^2) \to 0$. Then for every fixed integer $r > 0$,

$$\frac{\sum_{i=1}^r a_{i,i}^2}{\sum_{i=1}^n a_{i,i}^2} \le \frac{r \max_{1\le i\le n} a_{i,i}^2}{\sum_{i=1}^n a_{i,i}^2} \to 0.$$

Then there exists a sequence of positive integers $r(n) \to \infty$ such that $\left(\sum_{i=1}^{r(n)} a_{i,i}^2\right)/\left(\sum_{i=1}^n a_{i,i}^2\right) \to 0$ and $r(n)/n \to 0$. Write

$$Z^\top \mathbf{A} Z + b^\top Z - \operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{r(n)} a_{i,i}(z_i^2 - 1) + \sum_{i=1}^{r(n)} b_i z_i + \sum_{i=r(n)+1}^{n} \left(a_{i,i}(z_i^2-1) + b_i z_i\right),$$

which is a sum of independent random variables. The first term is negligible since $\operatorname{Var}(\sum_{i=1}^{r(n)} a_{i,i}(z_i^2 - 1)) = o(\sum_{i=1}^n a_{i,i}^2)$. Now we deal with the third term. From Berry-Esseen inequality (See, e.g., DasGupta (2008), Theorem 11.2), there exists an absolute constant $C^* > 0$, such that

$$\sup_{x\in\mathbb{R}} \left| \Pr\left(\frac{\sum_{i=r(n)+1}^n \left(a_{i,i}(z_i^2-1) + b_i z_i\right)}{\sqrt{2\sum_{i=r(n)+1}^n a_{i,i}^2 + \sum_{i=r(n)+1}^n b_i^2}} \le x \right) - \Phi(x) \right| \le C^* \frac{\sum_{i=r(n)+1}^n \operatorname{E}\left|a_{i,i}(z_i^2-1) + b_i z_i\right|^3}{\left(2\sum_{i=r(n)+1}^n a_{i,i}^2 + \sum_{i=r(n)+1}^n b_i^2\right)^{3/2}}.$$

By some simple algebra, there exist absolute constants $C_1^*, C_2^* > 0$ such that for sufficiently large n,

$$\sup_{x\in\mathbb{R}} \left| \Pr\left(\frac{\sum_{i=r(n)+1}^n \left(a_{i,i}(z_i^2-1) + b_i z_i\right)}{\sqrt{2\sum_{i=r(n)+1}^n a_{i,i}^2 + \sum_{i=r(n)+1}^n b_i^2}} \le x \right) - \Phi(x) \right| \le C_1^* \frac{\max_{1\le i\le n} |a_{i,i}|}{\sqrt{\sum_{i=1}^n a_{i,i}^2}} + C_2^* \frac{|b_{r(n)+1}|}{\|\mathbf{b}\|}.$$

Since the right hand side tends to 0, we have

$$\frac{\sum_{i=r(n)+1}^n \left(a_{i,i}(z_i^2-1) + b_i z_i\right)}{\sqrt{2\sum_{i=r(n)+1}^n a_{i,i}^2 + \sum_{i=r(n)+1}^n b_i^2}} \rightsquigarrow \mathcal{N}(0,1).$$

Note that $\sum_{i=r(n)+1}^n \left(a_{i,i}(z_i^2-1) + b_i z_i\right)$ is independent of $\sum_{i=1}^{r(n)} b_i z_i$ and $\sum_{i=1}^{r(n)} b_i z_i \sim \mathcal{N}(0, \sum_{i=1}^{r(n)} b_i^2)$. Thus,

$$\frac{\sum_{i=1}^{r(n)} b_i z_i + \sum_{i=r(n)+1}^n \left(a_{i,i}(z_i^2-1) + b_i z_i\right)}{\sqrt{2\sum_{i=1}^n a_{i,i}^2 + \sum_{i=1}^n b_i^2}} \rightsquigarrow \mathcal{N}(0,1).$$

This completes the proof.

□

Note that under the normality, $T_n - \operatorname{tr}((\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1})/(n-q)$ has zero mean.

Proof of Theorem 3. Note that $\mathbf{X}_b^* \mathbf{X}_b^{*\top} \sim \operatorname{Wishart}(p, \sigma_b^2 \mathbf{I}_{n-q})$.

$$\frac{\mathbf{y}^{*\top}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k \mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*} - \frac{\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k}{n-q} = \frac{\mathbf{y}^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k}{n-q}\mathbf{I}_{n-q}\right)\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*}.$$

We have

$$
\begin{aligned}
\mathbf{y}^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k}{n-q}\mathbf{I}_{n-q}\right)\mathbf{y}^* =& \boldsymbol{\varepsilon}^\top\tilde{\mathbf{U}}_a\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k}{n-q}\mathbf{I}_{n-q}\right)\tilde{\mathbf{U}}_a^\top\boldsymbol{\varepsilon} \\
&+ 2\boldsymbol{\varepsilon}^\top\tilde{\mathbf{U}}_a\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k}{n-q}\mathbf{I}_{n-q}\right)\mathbf{X}_b^*\boldsymbol{\beta}_b \\
&+ \boldsymbol{\beta}_b^\top\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k}{n-q}\mathbf{I}_{n-q}\right)\mathbf{X}_b^*\boldsymbol{\beta}_b \\
=&: A_1 + A_2 + A_3.
\end{aligned}
$$

To deal with $A_1 + A_2$, note that $\tilde{\mathbf{U}}_a^\top\boldsymbol{\varepsilon} \sim \mathcal{N}_{n-q}(0, \phi^{-1}\mathbf{I}_{n-q})$. We would like to apply Lemma 1 conditioning on $\mathbf{X}_b$. From lemma 3,

$$
\begin{aligned}
\operatorname{tr}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k}{n-q}\mathbf{I}_{n-q}\right)^2 =& (1+o_P(1))\sigma_b^{4k}p^{2k}(n-q)\operatorname{Var}(\eta^k), \\
\operatorname{tr}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k}{n-q}\mathbf{I}_{n-q}\right)^4 =& (1+o_P(1))\sigma_b^{8k}p^{4k}(n-q)\operatorname{E}(\eta^k - \operatorname{E}\eta^k)^4,
\end{aligned}
\tag{15}
$$

where η is a random variable with density function

$$
p_c(x) = \mathbf{1}_{\left[(1-c^{-1/2})^2,(1+c^{-1/2})^2\right]}(x)\frac{c}{2\pi x}\sqrt{4/c - (x-(1/c+1))^2}.
$$

It follows that

$$
\frac{\operatorname{tr}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k}{n-q}\mathbf{I}_{n-q}\right)^4}{\operatorname{tr}^2\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k}{n-q}\mathbf{I}_{n-q}\right)^2} \to 0.
$$

Then Lemma 1 implies that

$$
\frac{A_1 + A_2}{\sqrt{\operatorname{Var}(A_1|\mathbf{X}_b) + \operatorname{Var}(A_2|\mathbf{X}_b)}} \rightsquigarrow \mathcal{N}(0,1). \tag{16}
$$

We have

$$
\begin{aligned}
\operatorname{Var}(A_1|\mathbf{X}_b) &= 2\phi^{-2}\operatorname{tr}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k}{n-q}\mathbf{I}_{n-q}\right)^2 \\
&= (1+o_P(1))2\phi^{-2}\sigma_b^{4k}p^{2k}(n-q)\operatorname{Var}(\eta^k),
\end{aligned}
\tag{17}
$$

where the last equality follows from (15).

Now we deal with $\operatorname{Var}(A_2|\mathbf{X}_b)$. We have

$$
\operatorname{Var}(A_2|\mathbf{X}_b) = 4\phi^{-1}\boldsymbol{\beta}_b^\top\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k)}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\boldsymbol{\beta}_b.
$$

Let $\mathbf{O}$ be a $p \times p$ random matrix which is generated from Haar distribution and is independent of $\mathbf{X}_b$. The rotation invariance of normal distribution implies that $\mathbf{X}_b\mathbf{O}$ has the same distribution as

$\mathbf{X}_b$ and is independent of $\mathbf{O}$. Then

$$\begin{aligned}\operatorname{Var}(A_2|\mathbf{X}_b) &= 4\phi^{-1}\boldsymbol{\beta}_b^\top\mathbf{O}\mathbf{O}^\top\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{O}\mathbf{O}^\top\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{O}\mathbf{O}^\top\mathbf{X}_b^{*\top})^k)}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\mathbf{O}\mathbf{O}^\top\boldsymbol{\beta}_b\\ &\overset{d}{=} 4\phi^{-1}\boldsymbol{\beta}_b^\top\mathbf{O}\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k)}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\mathbf{O}^\top\boldsymbol{\beta}_b.\end{aligned}$$

Note that $\mathbf{O}^\top\boldsymbol{\beta}_b/\|\mathbf{O}^\top\boldsymbol{\beta}_b\|$ is uniformly distributed on the unit sphere S^{p-1}. From Lemma 2 and Lemma 3, we have

$$\begin{aligned}&\operatorname{E}\left(4\phi^{-1}\boldsymbol{\beta}_b^\top\mathbf{O}\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k)}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\mathbf{O}^\top\boldsymbol{\beta}_b\middle|\mathbf{X}_b\right)\\ =&4\phi^{-1}p^{-1}\|\boldsymbol{\beta}_b\|^2\operatorname{tr}\left(\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k)}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\right)\\ =&(1+o_P(1))4\phi^{-1}\|\boldsymbol{\beta}_b\|^2\sigma_b^{4k+2}p^{2k}(n-q)\operatorname{E}[(\eta^k-\operatorname{E}\eta^k)^2\eta],\end{aligned}$$

and

$$\begin{aligned}&\operatorname{Var}\left(4\phi^{-1}\boldsymbol{\beta}_b^\top\mathbf{O}\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k)}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\mathbf{O}^\top\boldsymbol{\beta}_b\middle|\mathbf{X}_b\right)\\ \le&\frac{32}{p^2}\phi^{-2}\|\boldsymbol{\beta}_b\|^4\operatorname{tr}\left(\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k)}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\right)^2\\ =&(1+o_P(1))32\phi^{-2}\|\boldsymbol{\beta}_b\|^4\sigma_b^{8k+4}p^{4k}(n-q)\operatorname{E}[(\eta^k-\operatorname{E}\eta^k)^4\eta^2].\end{aligned}$$

It follows that

$$\operatorname{Var}(A_2|\mathbf{X}_b) = (1+o_P(1))4\phi^{-1}\|\boldsymbol{\beta}_b\|^2\sigma_b^{4k+2}p^{2k}(n-q)\operatorname{E}[(\eta^k-\operatorname{E}\eta^k)^2\eta]. \tag{18}$$

Combining (16), (17) and (18) yields

$$\frac{A_1+A_2}{\sqrt{2\phi^{-1}\sigma_b^{4k}p^{2k}(n-q)\left(\phi^{-1}\operatorname{E}(\eta^k-\operatorname{E}\eta^k)^2+2\|\boldsymbol{\beta}_b\|^2\sigma_b^2\operatorname{E}[(\eta^k-\operatorname{E}\eta^k)^2\eta]\right)}} \rightsquigarrow \mathcal{N}(0,1). \tag{19}$$

Now we deal with $\mathbf{y}^{*\top}\mathbf{y}^*$. Note that

$$\begin{aligned}\operatorname{E}(\mathbf{y}^{*\top}\mathbf{y}^*|\mathbf{X}_b) =&\boldsymbol{\beta}_b^\top\mathbf{X}_b^{*\top}\mathbf{X}_b^*\boldsymbol{\beta}_b + \phi^{-1}(n-q),\\ \operatorname{Var}(\mathbf{y}^{*\top}\mathbf{y}^*|\mathbf{X}_b) \le&2\operatorname{Var}\left(2\boldsymbol{\varepsilon}^\top\tilde{\mathbf{U}}_a\mathbf{X}_b^*\boldsymbol{\beta}_b\right) + 2\operatorname{Var}\left(\boldsymbol{\varepsilon}^\top\tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top\boldsymbol{\varepsilon}\right)\\ =&O\left(\boldsymbol{\beta}_b^\top\mathbf{X}_b^{*\top}\mathbf{X}_b^*\boldsymbol{\beta}_b + n-q\right)\\ =&o\left(\left(\operatorname{E}(\mathbf{y}^{*\top}\mathbf{y}^*|\mathbf{X}_b)\right)^2\right).\end{aligned}$$

Consequently,

$$\mathbf{y}^{*\top}\mathbf{y}^* = (1+o_P(1))\left(\boldsymbol{\beta}_b^\top\mathbf{X}_b^{*\top}\mathbf{X}_b^*\boldsymbol{\beta}_b + \phi^{-1}(n-q)\right).$$

By the same techniques as we deal with $\operatorname{Var}(A_2|\mathbf{X}_b)$,

$$\mathbf{y}^{*\top}\mathbf{y}^* = (1+o_P(1))(n-q)\left(\sigma_b^2\|\boldsymbol{\beta}_b\|^2 \operatorname{E}\eta + \phi^{-1}\right). \tag{20}$$

It follows from (19) and (20) that

$$\frac{(n-q)\left(\sigma_b^2\|\boldsymbol{\beta}_b\|^2 \operatorname{E}\eta + \phi^{-1}\right)}{\sqrt{2\phi^{-1}\sigma_b^{4k}p^{2k}(n-q)\left(\phi^{-1}\operatorname{E}(\eta^k-\operatorname{E}\eta^k)^2 + 2\|\boldsymbol{\beta}_b\|^2\sigma_b^2\operatorname{E}[(\eta^k-\operatorname{E}\eta^k)^2\eta]\right)}}\frac{A_1+A_2}{\mathbf{y}^{*\top}\mathbf{y}^*} \rightsquigarrow \mathcal{N}(0,1). \tag{21}$$

Again by the same techniques as we deal with $\operatorname{Var}(A_2|\mathbf{X}_b)$,

$$A_3 = (1+o_P(1))\|\boldsymbol{\beta}_b\|^2\sigma_b^{2k+2}p^k(n-q)\left(\operatorname{E}\eta^{k+1} - \operatorname{E}\eta^k\operatorname{E}\eta\right).$$

Thus,

$$\begin{aligned}
&\frac{(n-q)\left(\sigma_b^2\|\boldsymbol{\beta}_b\|^2 \operatorname{E}\eta + \phi^{-1}\right)}{\sqrt{2\phi^{-1}\sigma_b^{4k}p^{2k}(n-q)\left(\phi^{-1}\operatorname{E}(\eta^k-\operatorname{E}\eta^k)^2 + 2\|\boldsymbol{\beta}_b\|^2\sigma_b^2\operatorname{E}[(\eta^k-\operatorname{E}\eta^k)^2\eta]\right)}}\frac{A_3}{\mathbf{y}^{*\top}\mathbf{y}^*}\\
=&\frac{(1+o_P(1))\sqrt{n-q}\|\boldsymbol{\beta}_b\|^2\sigma_b^2\left(\operatorname{E}\eta^{k+1}-\operatorname{E}\eta^k\operatorname{E}\eta\right)}{\sqrt{2\phi^{-2}\operatorname{E}(\eta^k-\operatorname{E}\eta^k)^2 + 4\phi^{-1}\|\boldsymbol{\beta}_b\|^2\sigma_b^2\operatorname{E}[(\eta^k-\operatorname{E}\eta^k)^2\eta]}}.
\end{aligned} \tag{22}$$

By a standard subsequence argument, we can without loss of generality and assume $\sqrt{n-q}\|\boldsymbol{\beta}_b\|^2 \to s \in [0,+\infty]$. If $s=+\infty$, the right hand side of (22) tends to infinity and the conclusion follows immediately. In what follows, we assume $s<+\infty$. In this case, since $\|\boldsymbol{\beta}_b\|^2 \to 0$, (21) becomes

$$\frac{\sqrt{n-q}}{\sqrt{2\sigma_b^{4k}p^{2k}\operatorname{E}(\eta^k-\operatorname{E}\eta^k)^2}}\frac{A_1+A_2}{\mathbf{y}^{*\top}\mathbf{y}^*} \rightsquigarrow \mathcal{N}(0,1).$$

And (22) becomes

$$\frac{\sqrt{n-q}}{\sqrt{2\sigma_b^{4k}p^{2k}\operatorname{E}(\eta^k-\operatorname{E}\eta^k)^2}}\frac{A_3}{\mathbf{y}^{*\top}\mathbf{y}^*} = \frac{\sqrt{n-q}\|\boldsymbol{\beta}_b\|^2\phi\sigma_b^2\left(\operatorname{E}\eta^{k+1}-\operatorname{E}\eta^k\operatorname{E}\eta\right)}{\sqrt{2\operatorname{E}(\eta^k-\operatorname{E}\eta^k)^2}} + o_P(1).$$

Thus,

$$\frac{\sqrt{n-q}}{\sqrt{2\sigma_b^{4k}p^{2k}\operatorname{E}(\eta^k-\operatorname{E}\eta^k)^2}}\left(R_k - \|\boldsymbol{\beta}_b\|^2\phi\sigma_b^{2k+2}p^k\left(\operatorname{E}\eta^{k+1}-\operatorname{E}\eta^k\operatorname{E}\eta\right)\right) \rightsquigarrow \mathcal{N}(0,1).$$

Then the conclusion follows.

□

Lemma 2. *Let* $\mathbf{A}$ *be an* $p\times p$ *symmetric matrix. Let* $Z=(z_1,\ldots,z_p)^\top$ *be a* p *dimensional random vector with uniform distribution on the unit sphere* S^{p-1}*. Then*

$$\operatorname{E}(Z^\top\mathbf{A}Z) = \frac{1}{p}\operatorname{tr}(\mathbf{A}),\quad \operatorname{Var}(Z^\top\mathbf{A}Z) = \frac{2}{p(p+2)}\left(\operatorname{tr}(\mathbf{A}^2) - \frac{1}{p}\operatorname{tr}^2(\mathbf{A})\right) \le \frac{2}{p^2}\operatorname{tr}(\mathbf{A}^2).$$

Proof. The result follows from direct calculation and the fact that for nonnegative integers $m_1, \dots, m_p$,

$$\mathrm{E}\prod_{i=1}^{p} z_i^{2m_i} = \frac{\Gamma(p/2)\prod_{i=1}^{p}\Gamma(m_i+1/2)}{\pi^{p/2}\Gamma(\sum_{i=1}^{p} m_i + p/2)}.$$

□

Let η be a random variable with density function

$$p_c(x) = \mathbf{1}_{[(1-c^{-1/2})^2,(1+c^{-1/2})^2]}(x)\frac{c}{2\pi x}\sqrt{4/c-(x-(1/c+1))^2}.$$

Define $\nu_{r,c} = \mathrm{E}\,\eta^r$.

The following lemma is a direct consequence of MP law and Bai Yin law.

Lemma 3. *Under the assumptions of Theorem 3, for every $r \in \mathbb{R}$,*

$$\frac{1}{\sigma_b^{2r} p^r (n-q)} \operatorname{tr}(\mathbf{X}_b^* \mathbf{X}_b^{*\top})^r \xrightarrow{a.s.} \nu_{r,c}$$

References

Bai, Z., Pan, G., and Yin, Y. (2018). A central limit theorem for sums of functions of residuals in a high-dimensional regression model with an application to variance homoscedasticity test. *TEST*, 27(4):896–920.

Bentkus, V. and Gotze, F. (1996). Optimal rates of convergence in the clt for quadratic forms. *Annals of Probability*, 24(1):466–490.

Buhlmann, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli*, 19(4):1212–1242.

Casella, G. and Moreno, E. (2006). Objective bayesian variable selection. *Journal of the American Statistical Association*, 101(473):157–167.

Chatterjee, S. (2008). A new method of normal approximation. *The Annals of Probability*, 36(4):1584–1610.

Chen, S. X., Zhang, L., and Zhong, P. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819.

Cohn, D. L. (2013). *Measure Theory.* Birkhauser Advanced Texts Basler Lehrbucher. Birkhuser Basel, 2 edition.

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability.* Springer Texts in Statistics. Springer-Verlag New York, 1 edition.

Dicker, L. H. and Erdogdu, M. A. (2017). Flexible results for quadratic forms with applications to variance components estimation. *The Annals of Statistics*, 45(1):386–414.

E. L. Lehmann, J. P. R. (2005). *Testing Statistical Hypotheses*. Springer New York.

Feng, L., Zou, C., Wang, Z., and Chen, B. (2013). Rank-based score tests for high-dimensional regression coefficients. *Electronic Journal of Statistics*, 7(0):2131–2149.

Goddard, S. D. and Johnson, V. E. (2016). Restricted most powerful bayesian tests for linear models. *Scandinavian Journal of Statistics*, 43(4):1162–1177.

Goeman, J. J., van de Geer, S. A., and van Houwelingen, H. C. (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):477–493.

Goeman, J. J., van Houwelingen, H. C., and Finos, L. (2011). Testing against a high-dimensional alternative in the generalized linear model: asymptotic type i error control. *Biometrika*, 98(2):381–390.

Götze, F. and Tikhomirov, A. (2002). Asymptotic distribution of quadratic forms and applications. *Journal of Theoretical Probability*, 15(2):423–475.

Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis*. Cambridge University Press.

Javier Girón, F., Lina Martínez, M., Moreno, E., and Torres, F. (2006). Objective testing procedures in linear models: Calibration of the p-values. *Scandinavian Journal of Statistics*, 33(4):765–784.

Jiang, J. (1996). Reml estimation: asymptotic behavior and related topics. *Annals of Statistics*, 24(1):255–286.

Lan, W., Wang, H., and Tsai, C. L. (2014). Testing covariates in high-dimensional regression. *Annals of the Institute of Statistical Mathematics*, 66(2):279–301.

Lan, W., Zhong, P.-S., Li, R., Wang, H., and Tsai, C.-L. (2016). Testing a single regression coefficient in high dimensional linear models. *Journal of Econometrics*, 195(1):154–168.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423.

Pollard, D. (1984). Convergence of stochastic processes. *Economica*.

Sevast'yanov, B. A. (1961). A class of limit distributions for quadratic forms of stochastic random variables. *Theory of Probability & Its Applications*, 6(3):337–340.

Wang, S. and Cui, H. (2015). A new test for part of high dimensional regression coefficients. *Journal of Multivariate Analysis*, 137:187–203.

Xianyang Zhang, G. C. (2017). Simultaneous inference for high-dimensional linear models. *Journal of the American Statistical Association*, 112(518):757–768.

Xu, K. (2016). A new nonparametric test for high-dimensional regression coefficients. *Journal of Statistical Computation and Simulation*, 87(5):855–869.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242.

Zhong, P. S. and Chen, S. X. (2011). Tests for high-dimensional regression coefficients with factorial designs. *Journal of the American Statistical Association*, 106(493):260–274.

Zhou, Q. and Guan, Y. (2018). On the null distribution of bayes factors in linear regression. *Journal of the American Statistical Association*, 0(0):1–10.